Naval Research Laboratory

Washington, DC 20375-5320

NRL/FR/5550--99-9921

Transcoding Between Two DoD Narrowband Voice Encoding Algorithms (LPC-10 and MELP)

GEORGE S. KANG
DAVID A. HEIDE

Transmission Technology Branch
Information Technology Division

October 15, 1999

The following notice applies to any unclassified (including originally classified and now declassified) technical reports released to "qualified U.S. contractors" under the provisions of DOD Directive 5230.25, Withholding of Unclassified Technical Data From Public Disclosure.

NOTICE TO ACCOMPANY THE DISSEMINATION OF EXPORT-CONTROLED TECHNICAL DATA

1. Export of information contained herein, which includes, in some circumstances, release to foreign nationals within the United States, without first obtaining approval or license form the Department of State for items controlled by the International Traffic in Arms Regulations (ITAR), or the Department of Commerce for items controlled by the Export Administration Regulations (EAR), may constitute a violation of law.

2. Under 22 U.S.C. 2778 the penalty for unlawful export of items or information controlled under the ITAR is up to two years imprisonment, or a fine of $100,000, or both. Under 50 U.S.C., Appendix 2410, the penalty for unlawful export of items or information controlled under the EAR is a fine of up to $1,000,000, or five times the value of the exports, whichever is greater; or for an individual, imprisonment of up to 10 years, or a fine of up to 10 years, or a fine of up to $250,000, or both.

3. In accordance with your certification that establishes you as a "qualified U.S. Contractor", unauthorized dissemination of this information is prohibited and may result in disqualification as a qualified U.S. contractor, and may be considered in determining your eligibility for future contracts with the Department of Defense.

4. The U.S. Government assumes no liability for direct patent infringement, or contributory patent infringement or misuse of technical data.

5. The U.S. Government does not warrant the adequacy, accuracy, currency, or completeness of the technical data.

6. The U.S. Government assumes no liability for loss, damage. Or injury resulting form manufacture or use for any purpose of any product, article, system, or material involving reliance upon any or all technical data furnished in response to the request for technical data.

7. If the technical data furnished by the Government will be used for commercial manufacturing or other profit potential, a license for such use may be necessary. Any payments made in support of the request for data do not include or involve any license rights.

8. A copy of this notice shall be provided with any partial or complete reproduction of these data that are provided to qualified U.S. contractors.

DESTRUCTION NOTICE

For classified documents, follow the procedures in DOD 5200.22-M, Industrial Security Manual, Section II-19 or DOD 5200.1-R, Information Security Program Regulation, Chapter IX. For unclassified, limited documents, destroy by any method that will prevent disclosure of contents or reconstruction of the document.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
October 15, 1999	Continuing	01 Oct 98 – 30 July 1999

4. TITLE AND SUBTITLE
Transcoding Between Two DoD Narrowband Voice Encoding Algorithms (LPC-10 and MELP)

5. FUNDING NUMBERS
33904N
61153N

6. AUTHOR(S)
George S. Kang and David A. Heide

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/FR/5550--99-9921

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
Space and Naval Warfare Systems Command
4301 Pacific Highway
San Diego, CA 92110-3127

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

For nearly 20 years, DoD has had only one narrowband voice algorithm called Linear Predictive Coder (LPC). It is used in the Advanced Narrowband Digital Voice Terminals (ANDVTs) operating at 2400 bits per second (b/s). Currently, 40,000 ANDVTs have been deployed by the Navy, Army, Air Force, Marine Corps, and special government agencies. DoD is currently planning to develop a new narrowband voice terminal called the Future Narrowband Digital Terminal (FNBDT), which features a new voice processing algorithm called Mixed Excitation Linear Predictor (MELP) operating at 2400 b/s. In the future, LPC must interoperate with MELP. Therefore, it is essential to develop a technique that enables MELP and LPC to interoperate, so that secure voice service among narrowband users will not be interrupted during the transition period.

Although LPC and MELP could interoperate through the age-old tandeming method, resultant speech degradation would be very severe because the bit stream must be converted to the speech waveform, which is re-analyzed and re-encoded. Therefore, NRL investigated an alternative interoperation technique, called "transcoding," where speech parameters, such as pitch, amplitude parameters, and filter parameters, are directly converted from one to the other vocoder. This report documents the computational steps required for transcoding and their theoretical basis. According to formalized tests, transcoding did not degrade speech intelligibility in comparison with LPC-10.

14. SUBJECT TERMS
narrowband speech encoding MELP
Transcoding between LPC-10 and MELP

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UL	SAR

CONTENTS

TRANSCODING BETWEEN TWO DOD NARROWBAND VOICE ENCODING ALGORITHMS (LPC-10 AND MELP)

INTRODUCTION

Voice communication is indispensable in tactical environments where speedy and interactive exchange of information is vital for accomplishing the mission. A tactical voice rate must be such that data rate must be low for narrowband links, verbal messages must be delivered in real-time, received messages must be intelligible enough even in noisy listening environments, and speakers' emotional states must be perceivable through spoken messages. Most important, all tactical voice terminals must interoperate in order to accomplish efficiently the common mission among the forces.

Interoperability of narrowband tactical terminals has been no problem for many years because there has been only one narrowband secure voice terminal in operation — the Advanced Narrowband Digital Voice Terminal (ANDVT) (Fig. 1), first developed in the late 1970s and early 1980s [1]. Over the years, 40,000 ANDVTs have been deployed by the Navy, Army, Air Force, Marine Corps, and special government agencies. These ANDVTs operate at 2400 bits per second (b/s).

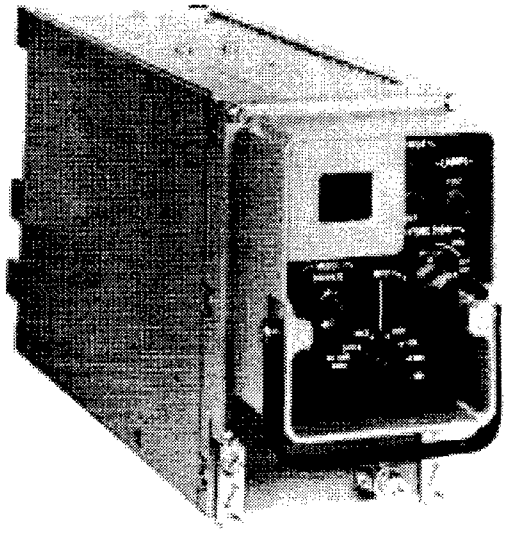


Fig. 1 — ANDVT, front-view. ANDVT combines voice processor, crypto, high frequency (HF), and line-of-sight (LOS) modems. ANDVT has three terminal configurations: (1) the tactical terminal (shown on the left) for shipboard, submarine, vehicular, tactical shelter, and airborne use, (2) the miniaturized terminal for man pack use, and (3) the airborne terminal for specifically airborne platforms. ANDVT was developed as a tri-service program with the Navy as a tactical agent. A notable feature of ANDVT is that it uses a common frame size of 22.5 ms (180 speech samples) for the voice processor, modem, and crypto to facilitate the aquisition and maintenance of synchronization. The voice processor features a 10-tap Linear Predictive Coder (LPC-10), which produced higher speech intelligibility and quality than existing channel vocoders. Over the years, there have been very few complaints about ANDVTs from the users.

After nearly 20 years of service, there is a need for a new narrowband voice terminal to meet future DoD requirements. In fact, DoD is currently planning to develop a new narrowband voice terminal called the Future Narrowband Digital Terminal (FNBDT) [2]. This terminal will use a new voice processing algorithm called Mixed Excitation Linear Predictor (MELP) operating at 2400 b/s [3] in a variety of networks.

Therefore, it is essential to develop a technique that enables the interoperation of FNBDT and ANDVTs as MELP is being deployed so that secure voice service among narrowband users will not be interrupted. We developed such a technique in this report. It is called *transcoding*, which directly converts the bit stream of

Manuscript approved August 9, 1999.

LPC-10 to the bit stream of MELP and vice versa. We envision that transcoding will be performed at a gateway located near the ANDVT or MELP receiver. Hence, interoperation does not require any modification for ANDVT nor any special design constraints for FNBDT.

Over the years, the interoperation between two different voice terminals was effected through *tandeming*. In the tandeming approach, one voice terminal generates the speech waveform, which in turn is re-analyzed and re-encoded by the second voice terminal. These re-analysis and re-encoding processes often introduce serious speech degradation. In contrast, *transcoding* converts speech parameters directly from one voice terminal to another. Hence, speech degradation is far less than what is expected from the tandeming approach.

The important and timely study documented in this report was sponsored by the Navy INFOSEC office (SPAWAR PMW161). They were not only the ANDVT technical agent during the developmental phase, but they are also a procurement agency of ANDVT. In addition, they are interested in the secure voice technology development aiming at higher speech quality, transparent security, and joint/allied interoperability. They were instrumental in developing the new voice processing techniques used in various government voice terminals, such as LPC-10 improvements used in STU-III (2400 b/s mode), line spectrum pairs (LSPs) used in STU-III (4800-b/s), the residual-excited LPC used in Motorola version of STU-III (9600-b/s), MELP error protection by ANDVT HF modem, and multirate processor (MRP) to integrate the narrowband and wideband voice resources into a single interoperable capability. The present transcoding study results will benefit especially the Navy because naval tactical voice communications are heavily dependent on narrowband channels.

BACKGROUND

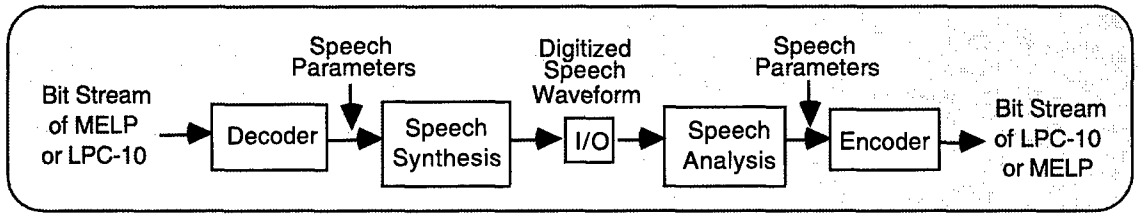
The interoperation of two different voice encoders requires the conversion of the bit streams from one encoder to the other. The old way was tandeming, and the new way is transcoding. We will give a brief overview for both approaches.

Tandeming

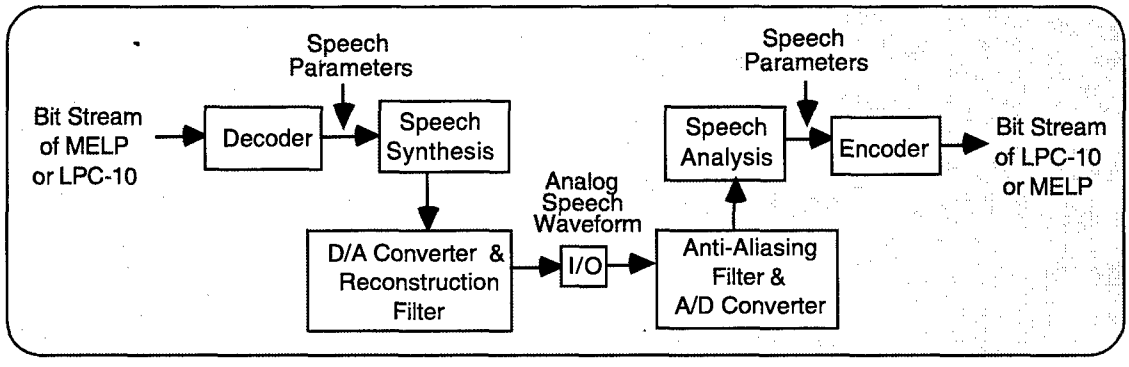
Tandeming is an age-old technique to interoperate two different voice encoders. As indicated in Fig. 2, the bit stream of one voice encoder is decoded to speech parameters (pitch, rms, filter coefficients, etc.). Then, the speech parameters are converted to the speech waveform. Finally, the speech waveform (in either analog or digitized form) is re-analyzed and re-encoded to become the bit stream of the tandeming voice encoder. Tandeming is essentially a back-to-back operation of two different voice encoders.

An advantage of tandeming is that any two vocoders (each with a different speech analysis principle, data rate, frame rate, etc.) can be linked to interoperate. A disadvantage is that speech is often degraded significantly due to a multitude of operations in the tandeming link, especially in analog tandeming (Fig. 2(b)) where two sets of anti-aliasing and reconstruction filters and A/D and D/A converters are present.

A tandem interface may be designed so that the speech waveform can be transferred in the digital form (Fig. 2(a)). A digital tandem interface eliminates D/A and A/D converters and reconstruction and anti-aliasing filters. As a result, speech will not be degraded as much. A digital tandem interface, however, must recognize each digitized speech waveform amplitude.



(a) Digital Tandeming



(b) Analog Tandeming

Fig. 2 — Tandem configuration. An important feature of tandeming is the regeneration of the speech waveform at the interface. Analog tandeming introduces significant speech degradation because the speech signal must be passed again through the D/A converter, reconstruction filter, anti-aliasing filter, and A/D converter.

Transcoding

Transcoding is not digital tandeming. Figure 3 shows that transcoding does not convert the incoming bit stream to the speech waveform. Rather, the incoming bit stream is converted to speech parameters, which are then converted to speech parameters of the interoperating voice encoder. An advantage of transcoding is that speech will not be degraded as much as tandeming because speech parameters are directly converted (not by the re-analysis of the speech waveform). A disadvantage of the transcoding approach is that the two interoperating voice encoders must be closely related, as with LPC-10 and MELP, to be discussed in the next section.

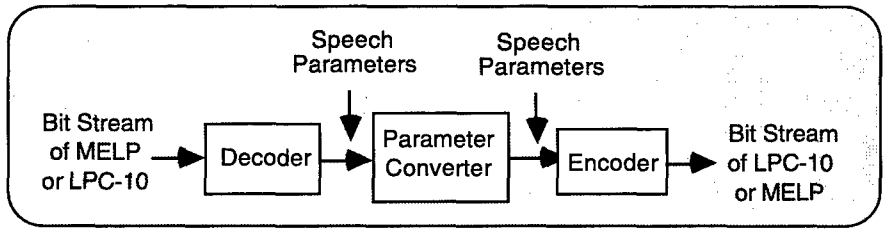
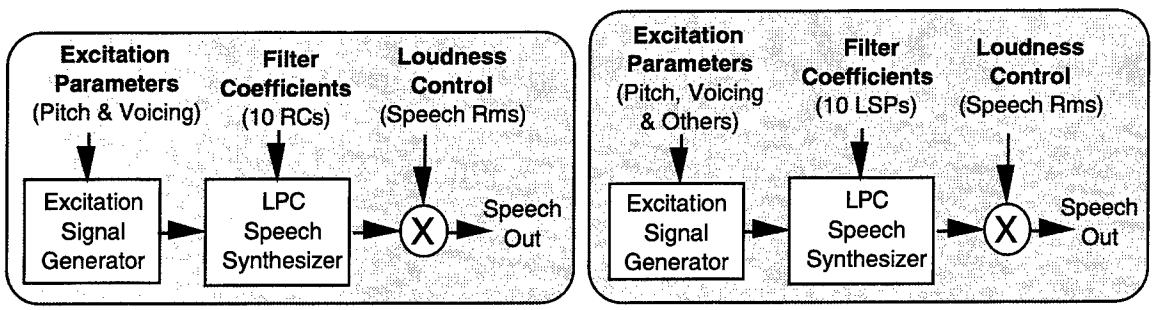


Fig. 3 — Transcoding process. A significant difference between transcoding and tandeming is that transcoding does not regenerate and re-analyze the speech waveform. Instead, speech parameters of one voice encoder are directly converted to speech parameters of the other voice encoder.

Speech Models for LPC-10 and MELP

Figure 4 shows that MELP and LPC-10 are closely related because they both use the identical speech analysis technique (i.e., linear predictive encoding), the identical frame size (180 samples), the identical speech sampling frequency (8 kHz), and the identical data rate (2400 b/s). Both use the same synthetic excitation signal (also known as the pitch excitation signal). Because of these similarities, transcoding is well-suited for MELP and LPC-10.



(a) LPC-10 (b) MELP

Fig. 4 — Speech generation models for LPC-10 and MELP. Both LPC-10 and MELP use an LPC-based speech synthesizer driven by a synthetically generated excitation signal source. A major difference between the LPC-10 and MELP speech models is that the MELP uses a more elaborate excitation signal. Details will be discussed in connection with transcoding of the excitation parameters. Because of basic similarities between the two, speech parameters (indicated by bold letters) can be converted directly from LPC-10 to MELP (and vice versa) without regenerating the speech waveform as required by tandeming. As will be discussed later, RCs and LSPs are the abbreviations for "reflection coefficients" and "line spectrum pairs," respectively.

Factor that Complicates Transcoding

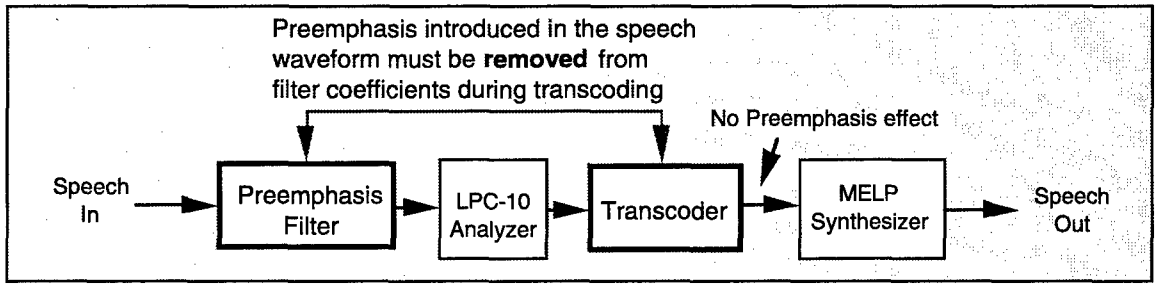
There is a factor that complicates transcoding between LPC-10 and MELP, however. Figure 5 shows that LPC-10 preemphasizes (i.e., boosts high frequencies and attenuates low frequencies) the speech waveform prior to the LPC analysis, whereas MELP does not. The presence or absence of preemphasis must be properly compensated during transcoding of both the speech root mean square (rms) parameter and filter coefficients. We will discuss the preemphasis compensation in detail.

Preemphasis Characteristics

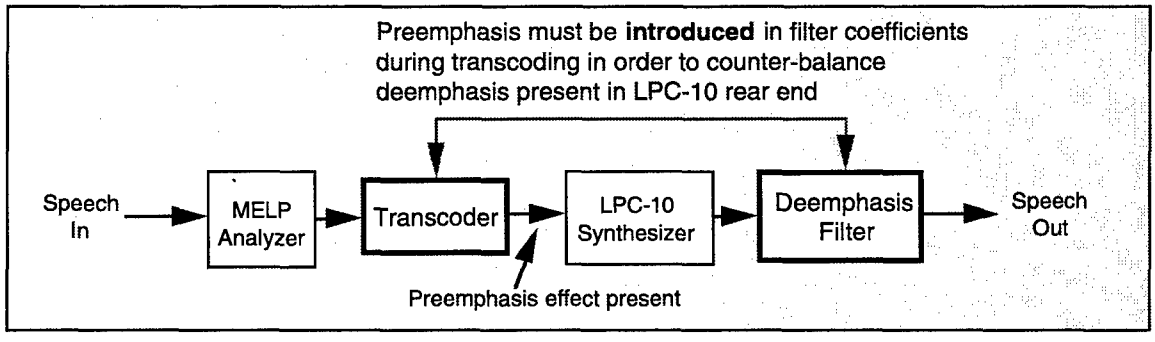
The purpose of preemphasizing the speech waveform is to reduce lower frequency components while boosting higher frequency components of the speech waveform. A digital filter adequate for preemphasis has a single zero. Such a preemphasis filter, denoted by $H_{PE}(z)$ has the transfer function

$$H_{PE}(z) = 1 - \frac{31}{32} z^{-1} . \tag{1}$$

This is the preemphasis filter specified in 1980 for ANDVT [1], and it was also specified in Federal Standard 1015 for the government-standard LPC-10 in 1984 [4].



(a) When LPC-10 is Interoperating with MELP



(b) When MELP is Interoperating with LPC-10

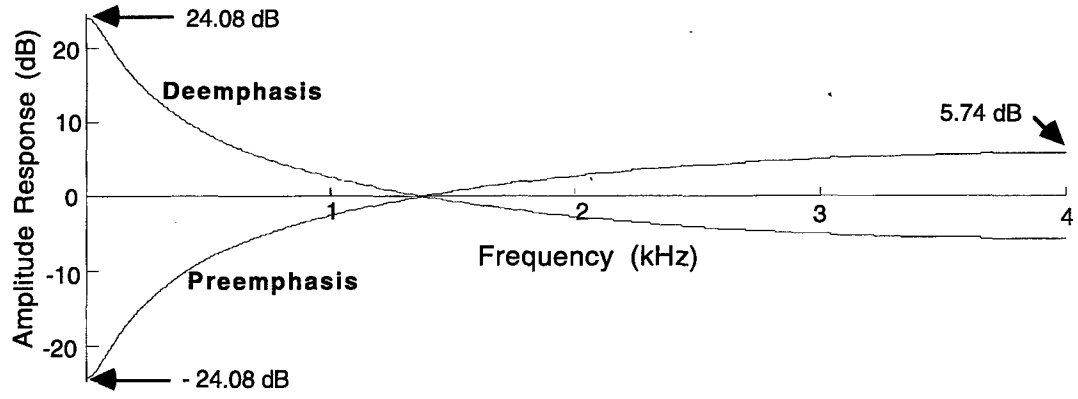
Fig. 5 — Presence of preemphasis in LPC-10 and absence of preemphasis in MELP. Due to a mismatch of preemphasis between LPC-10 and MELP, speech parameters, such as speech rms and filter coefficients, must be properly compensated when LPC-10 interoperates with MELP, and vice versa.

Once the speech waveform is preemphasized at the front end as in LPC-10, it is necessary to reverse the process (i.e., deemphasize) at the rear end to cancel the preemphasis. The transfer function of the deemphasis filter, denoted by $H_{DE}(z)$, is the inverse function of the transfer function of the preemphasis filter

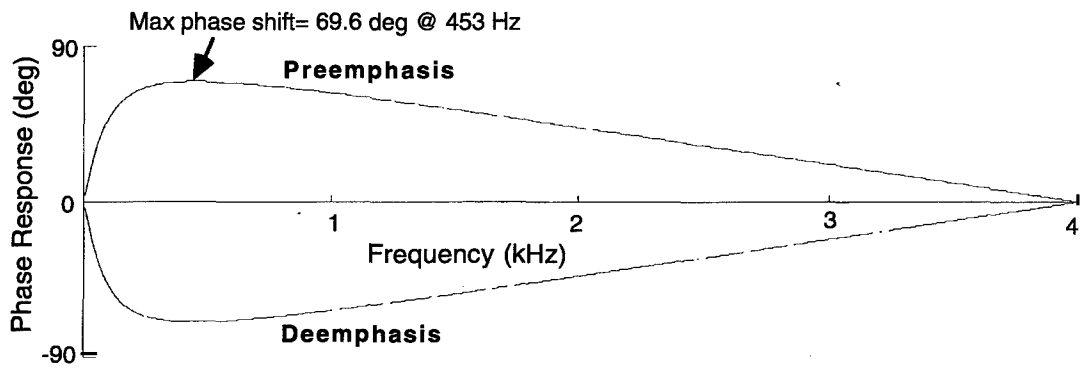
$$H_{DE}(z) = \frac{1}{1 - \frac{31}{32} z^{-1}} \cdot \tag{2}$$

Figure 6 shows frequency responses of both preemphasis and deemphasis filters.

Preemphasis has been used for the speech analysis/synthesis or voice encoding for many years. An advantage of preemphasizing the speech waveform prior to the analysis is to make the speech spectrum more balanced between lower and higher frequencies. As noted from the speech spectrum of a vowel shown in Fig. 7(a), low frequencies are strong and high frequencies are weak. If speech is too loud, lower frequencies are often clipped causing distortion. On the other hand, higher frequencies are often so weak that the speech analysis results are poor when representing or characterizing these frequency components. Preemphasis makes the speech spectrum more balanced between lower and higher frequencies, resulting in a spectral tilt that is less steep (Fig. 7(b)).



(a) Amplitude Response



(b) Phase Response

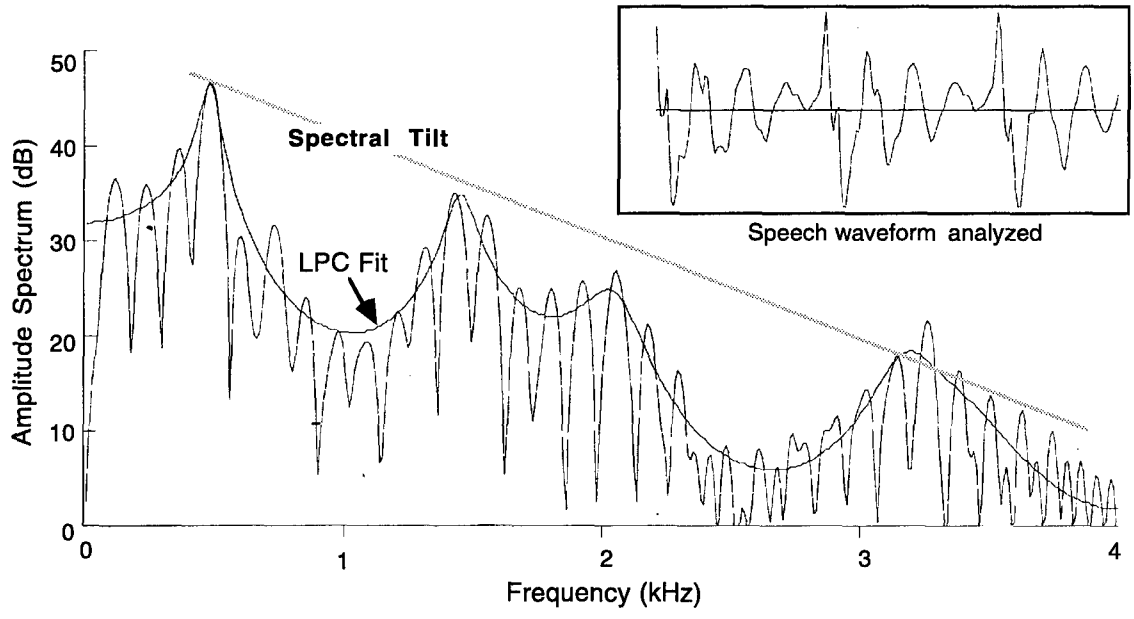
Fig. 6 — Frequency responses of the preemphasis and deemphasis filters used in LPC-10. As will be discussed, the amplitude response is essential for the rms transcoding, and the phase response plays a critical role in the filter coefficient transcoding.

TRANSCODING OF RMS PARAMETER

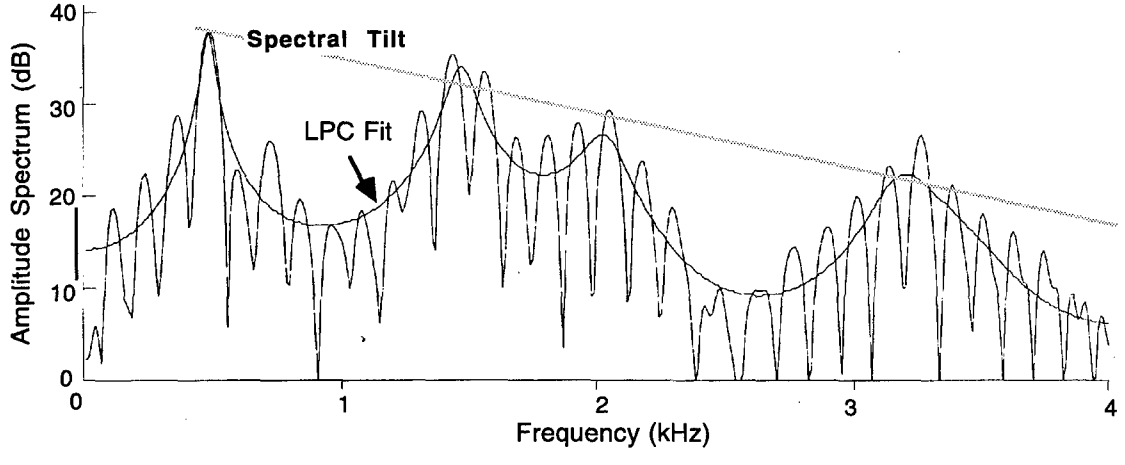
One of the speech parameters transmitted by LPC-10 and MELP is the root-mean-square (rms) value of the speech waveform, which controls the loudness of the synthesized speech (Fig. 4). The rms parameter, therefore, must be transcoded. As stated earlier, MELP computes the rms value of the original speech waveform, whereas LPC-10 computes the rms value of the *preemphasized* speech waveform. Therefore, transcoding of the rms parameters must include steps to compensate the presence and absence of the preemphasis.

Background

The difference between the LPC-10 rms and MELP rms is dependent on the speech spectrum in relation to the frequency response of the preemphasis filter response shown earlier in Fig. 6. The rms value of preemphasized speech (LPC-10) is generally smaller than the rms value of non-preemphasized speech (MELP), but they crisscross constantly. When the speech waveform has predominantly high frequencies (i.e., fricatives) the preemphasized rms value exceeds that of non-preemphasized rms value. Therefore, we should discard any notion for using a constant factor to convert the LPC-10 rms value to MELP rms value, and vice versa. Figure 8 illustrates a complex nature of the rms histograms for LPC-10 and MELP with the time-aligned speech spectrogram.

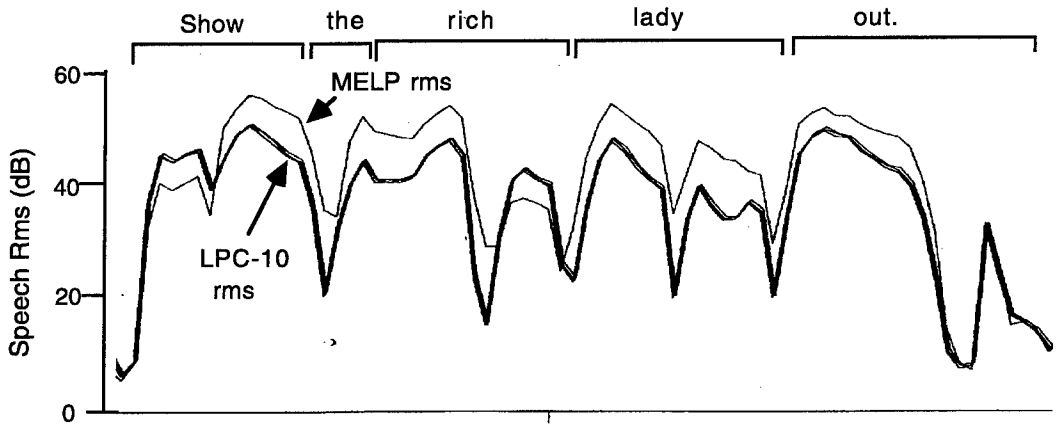


(a) Without Preemphasis. As noted, lower frequencies are much stronger than higher frequencies (i.e., the spectral tilt is rather steep).

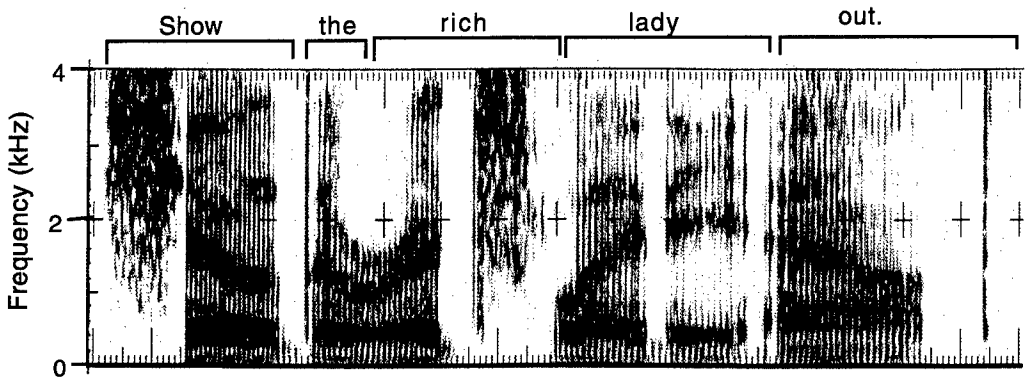


(b) With Preemphasis. The speech spectrum is more balanced between lower and higher frequencies. As a result, the spectral tilt is less steep.

Fig. 7 — Speech spectra without preemphasis (for MELP) and with preemphasis (for LPC-10). As noted in this figure, preemphasis reduces the magnitude of the spectral tilt. In other words, high- and low-frequency components are more equalized to produce an improved speech analysis result.



(a) Rms Histograms of LPC-10 and MELP



(b) Spectrogram of Original Speech

Fig. 8 — Rms histograms of MELP and LPC-10 and the time-aligned speech spectrogram. Figure 8(a) shows that the MELP rms is not proportional to LPC-10 rms. Therefore, the ratio of MELP rms to LPC-10 rms is not a constant. The MELP rms (thin line) is generally greater than LPC-10 rms (thick line) except for fricatives (such as /sh/, and /ch/ in this example), where high frequencies are dominant, as shown in Fig. 8(b).

Transcoding from LPC-10 Rms to MELP Rms

Figure 9 shows that transcoding of the rms between LPC-10 and MELP requires four steps.

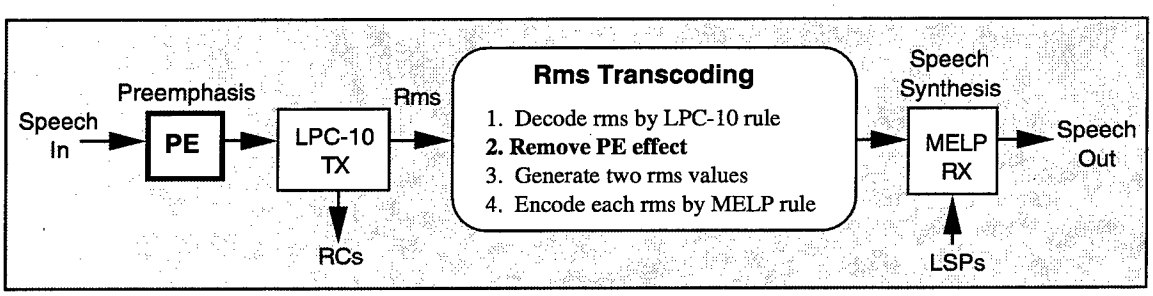


Fig. 9 — Steps required to transcode the rms parameter from LPC-10 to MELP. The most critical step is introduction or removal of the preemphasis (PE) effect in the rms value.

Steps 1 and 4 need not be elaborated because these rules are well defined and currently being used in LPC-10 and MELP.

Step 2: Remove Preemphasis Effect in Rms Value

Step 2 is critical in the rms transcoding from LPC-10 to MELP. To perform this step, we need either entire speech time samples or spectral samples. Unfortunately, we have neither of these in the bit stream. We have, however, the speech spectral envelope estimated from LPC coefficients (see Fig. 7 for an example). We use both speech spectral envelopes of LPC-10 and MELP for the rms transcoding. Step 2 is carried out in the following four stages:

(i) *RC-to-PC Conversion:* The reflection coefficients (RCs) from LPC-10 are converted to prediction coefficients (PCs). The well-known RC-to-PC conversion equation is given in most digital signal processing textbooks [5].

$$\beta_{j|n+1} = \beta_{j|n} - k_{n+1}\beta_{n+1-j|n} \qquad \text{j = 1, 2, ..., n} \tag{3}$$

with

$$\beta_{n+1|n+1} = k_{n+1}, \tag{4}$$

where $\beta_{j|n+1}$ means the jth prediction coefficient (with preemphasis) in the $(n+1)^{\text{th}}$ iteration.

(ii) *Compute speech envelope of preemphasized speech estimated by LPC-10:* Using the transformed PCs, the speech spectral envelope may be obtained from the basic LPC speech model

$$S_{LPC-10}(\omega) = \left. \frac{1}{1-\beta_1 z^{-1} - \beta_2 z^{-2} \ \ldots \ -\beta_{10} z^{-10}} \right|_{z=j\omega\tau}, \tag{5}$$

where βs are PCs transformed from RCs generated by LPC-10, τ is speech sampling time interval, and ω is frequency in rad/s. The speech spectral envelope estimated by LPC-10 is shown in Fig. 10 where the speech spectral envelope estimated by MELP is also shown for comparison.

(iii) *Compute speech envelope of non-preemphasized speech estimated by MELP:* Once the speech spectral envelope of the preemphasized case is known, the speech spectral envelope of non-preemphasized case can be obtained by the transformation utilizing the frequency response of the preemphasis filter. Thus,

$$S_{MELP}(\omega) = \frac{S_{LPC-10}(\omega)}{H_{PE}(\omega)}, \tag{6}$$

where $S_{MELP}(\omega)$ is the speech spectral envelope of MELP converted from the speech spectral envelope of LPC-10, $S_{LPC-10}(\omega)$. In Eq. (6), $H_{PE}(\omega)$ is the frequency response of the preemphasis filter defined by Eq. (1). Figure 10 illustrates speech spectral envelopes estimated by MELP where the speech envelope estimated by LPC-10 is also shown for comparison.

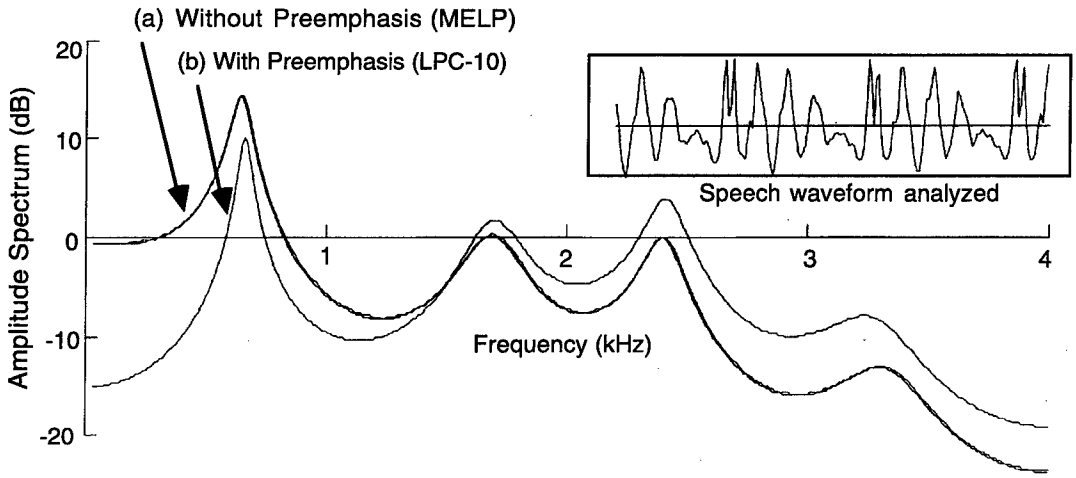


Fig. 10 — Speech spectral envelopes obtained from filter coefficients of LPC-10 or MELP. Note if the speech envelope with preemphasis is known, the speech spectral envelope without preemphasis (and vice versa) can be computed.

(iv) *Rms ratio between preemphasized and non-preemphasized speech:* Since the rms value of time samples equals the rms value of its spectral samples, the following relationship holds:

$$\frac{rms_{LPC-10}}{rms_{MELP}} = \frac{\sqrt{\left|\sum_{\omega=0}^{\Omega} S_{LPC-10}(\omega)\right|^2}}{\sqrt{\left|\sum_{\omega=0}^{\Omega} S_{MELP}(\omega)\right|^2}}. \tag{7}$$

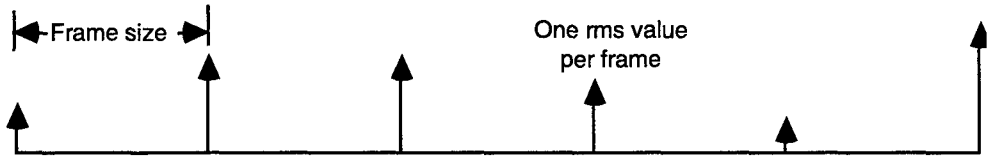
where Ω is the upper cutoff frequency of the speech signal (i.e., $2\pi(4000)$ radians). In Eq. (7), a 400 spectral summation from 0 to 4 kHz would be adequate. A 10 Hz frequency step is small enough to observe even a sharp resonant frequency.

This rms ratio is used as an rms correction factor between LPC-10 and MELP.

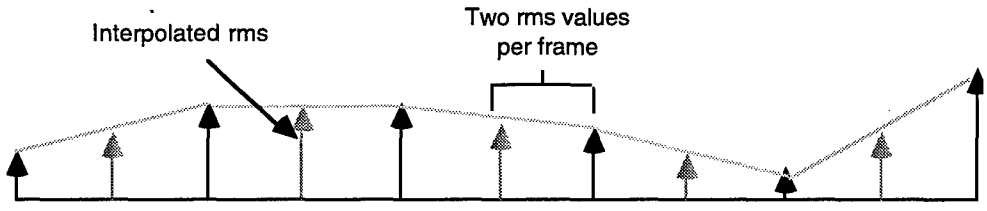
$$rms_{MELP} = \left(\frac{\sqrt{\left|\sum_{\omega=0}^{\Omega} S_{MELP}(\omega)\right|^2}}{\sqrt{\left|\sum_{\omega=0}^{\Omega} S_{LPC-10}(\omega)\right|^2}} \right) rms_{LPC-10}, \tag{8}$$

Step 3: Convert One Rms Value per Frame to Two Rms Values

LPC-10 transmits one rms value per frame, whereas MELP transmits two rms values per frame. Therefore, we must generate an additional rms value from LPC-10 in order to make a compatible bit stream with MELP (although an additional rms value doesn't improve speech). Figure 11 illustrates that this additional rms value is best generated by the computed rms at the midpoint of the frame through interpolation.



(a) LPC-10 Rms Generated Once Per Frame



(b) MELP Rms Regenerated from LPC-10 Rms Through Interpolation

Fig. 11 — LPC-10 rms and interpolated MELP rms. LPC-10 generates one rms value per frame, whereas MELP generates two rms values per frame. Therefore, when LPC-10 is interoperating with MELP, an intraframe rms value must be generated to make the converted rms bit stream compatible with the MELP's rms bit stream.

Demonstration of Rms Conversion Accuracy from LPC-10 to MELP

We illustrate the accuracy of the converted rms using Eq. (8). We performed the following operations and plotted the result:

- *MELP rms from the speech waveform (Goal):* The histogram of MELP rms is computed from the original speech waveform and plotted in Fig. 12 (thin line).
- *LPC-10 rms from the preemphasized speech waveform (Given):* The rms histogram of LPC-10 is also computed from the preemphasized speech waveform and plotted in Fig. 12 (thick line).
- *Transcoded rms for LPC-10 from MELP rms:* Based on Eq. (8), we converted the LPC-10 rms to the MELP rms. Results are plotted by cross marks in Fig. 12. As noted, cross marks are often right, indicating that the converted rms is rather accurate. Accuracy suffers only when speech is very soft (about -30 dB of the loudest). Rms errors in very soft speech are inconsequential because we can hardly hear them.

Transcoding from MELP Rms to LPC-10 Rms

Transcoding of the MELP rms to the LPC-10 rms is essentially a reverse process of that discussed in the preceding section. There are four steps in this rms transcoding process, as indicated in Fig. 13. Again, Steps 1 and 4 need no further elaboration because parameter encoding and decoding tables are well defined, and they have been implemented in current LPC-10 and MELP.

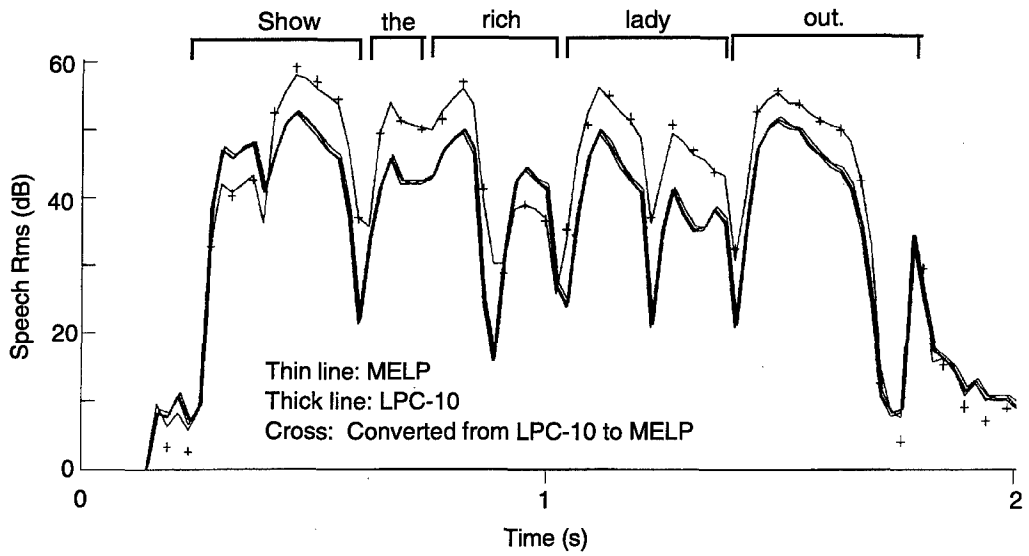


Fig. 12 — Rms histograms of LPC-10 (thick line), MELP (thin line), and converted results from LPC-10 to MELP (cross marks). The converted rms values from LPC-10 to MELP are in good agreement with the original MELP rms values. It may be rather hard to get any better results than this.

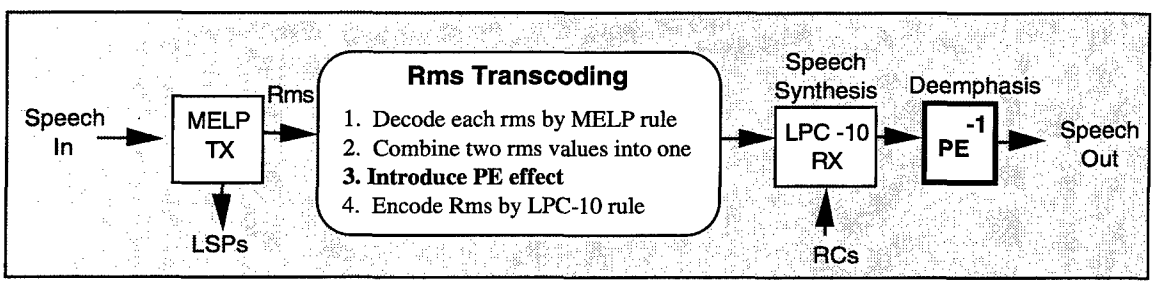


Fig. 13 — Steps required to transcode the rms parameter from MELP to LPC-10. As in transcoding of the rms value from LPC-10 to MELP, a critical step is introduction or removal of the preemphasis effect (PE) in the rms value.

Step 2: Combine Two Rms Values to One

Two incoming rms values per frame from MELP may be averaged to generate one rms value for LPC-10. Alternatively, the second rms value from MELP may be used as the LPC-10 rms value without averaging.

Step 3: Introduce the Preemphasis Effect in the Rms Value

All necessary processing equations have been derived in the preceding section for converting the MELP rms to LPC-10 rms. From Eq. (7), the LPC-10 rms in terms of the MELP rms is expressed by

$$rms_{LPC-10} = \left(\frac{\sqrt{\left| \sum_{\omega=0}^{\Omega} S_{LPC-10}(\omega) \right|^2}}{\sqrt{\left| \sum_{\omega=0}^{\Omega} S_{MELP}(\omega) \right|^2}} \right) rms_{MELP}, \tag{9}$$

where $S_{LPC-10}(\omega)$ is the speech spectral envelope of LPC-10 converted from that of MELP $S_{MELP}(\omega)$ by making the use of the relationship

$$S_{LPC-10}(\omega) = H_{PE}(\omega)S_{MELP}(\omega)\,, \tag{10}$$

where $H_{PE}(\omega)$ is the frequency response of the preemphasis filter shown in Fig. 5 earlier.

Demonstration of Rms Conversion Accuracy from MELP to LPC-10

We illustrate the accuracy of the converted rms using Eq. (9). We performed the following operations and plotted the results in Fig. 14:

- *MELP rms from the speech waveform (Given):* The histogram of MELP rms is computed from the original speech waveform and plotted in Fig. 14 (thin line).
- *LPC-10 rms from the preemphasized speech waveform (Goal):* The rms histogram of LPC-10 is also computed from the preemphasized speech waveform and plotted in Fig. 14 (thick line).
- *Transcoded rms for MELP from LPC-10 rms:* Based on Eq. (9), we converted the MELP rms to the LPC-10 rms. Results are plotted by cross marks in Fig. 14. As noted, cross marks are often right on the thick line, which indicates that the converted rms is rather accurate.

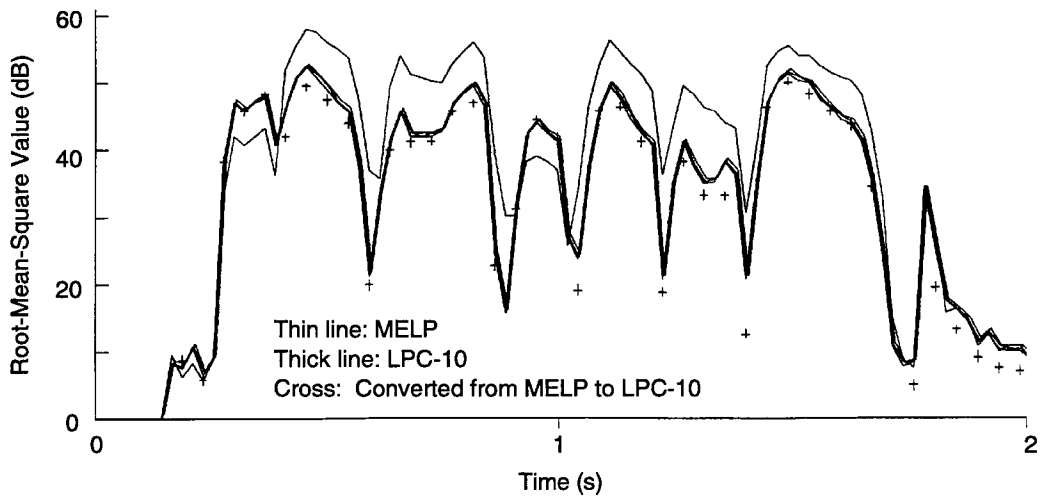


Fig. 14 — Rms histograms of LPC-10 (thick line), MELP (thin line), and converted results from MELP to LPC-10 (cross marks). Converted rms values are within a few dB, indicating the rms conversion algorithm is good.

TRANSCODING OF FILTER COEFFICIENTS

As indicated in Fig. 4, both LPC-10 and MELP transmit 10 filter coefficients. These filter coefficients add resonant frequencies to the spectrally white excitation signal so that the speech synthesizer output sounds like speech. Therefore, filter coefficients also must be transcoded as the rms parameter. LPC-10 converts prediction coefficients (PCs) to reflection coefficients (RCs) before transmission, whereas MELP converts PCs to line spectrum pairs (LSPs).

As will be discussed, transcoding of filter coefficients also includes a compensation for the preemphasis effect. In other words, preemphasis introduced in the speech waveform must be removed by filter coefficients when LPC-10 is interoperating with MELP. Conversely, when MELP is interoperating with LPC-10, the preemphasis effect must be introduced in filter coefficients because the LPC-10 trailing-end has a deemphasis filter to nullify the preemphasis. Transcoding of filter coefficients, therefore, is a major hurdle in the transcoding between MELP and LPC-10.

Background

Three Related Filter Coefficients

In both LPC-10 and MELP, the speech waveform is processed by the linear prediction analysis to generate PCs. As indicated in Fig. 15, PCs may be converted to RCs, which are transmitted by LPC-10, or LSPs, which are transmitted by MELP.

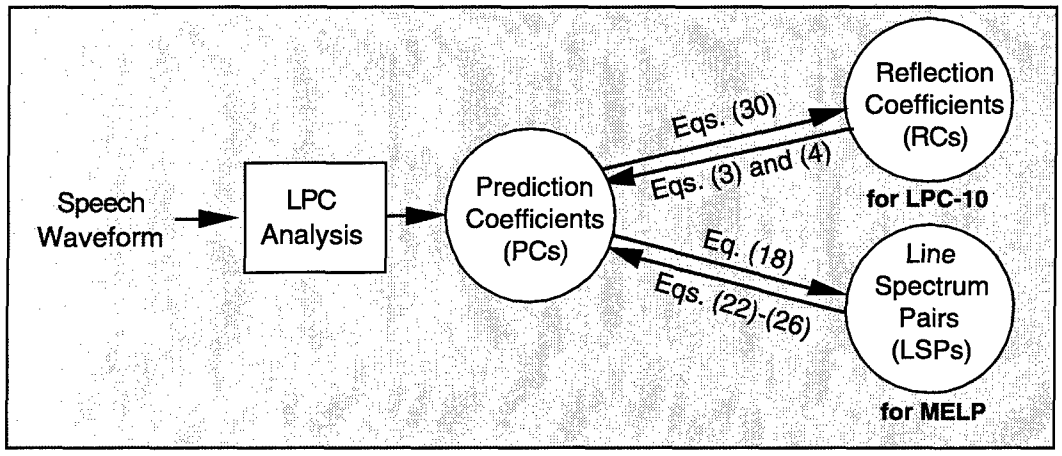


Fig. 15 — LPC coefficients. There are at least three different forms of LPC coefficients that are often used for speech spectral representation — prediction coefficients (PCs), reflection coefficients (RCs) and line-spectrum pairs (LSPs).

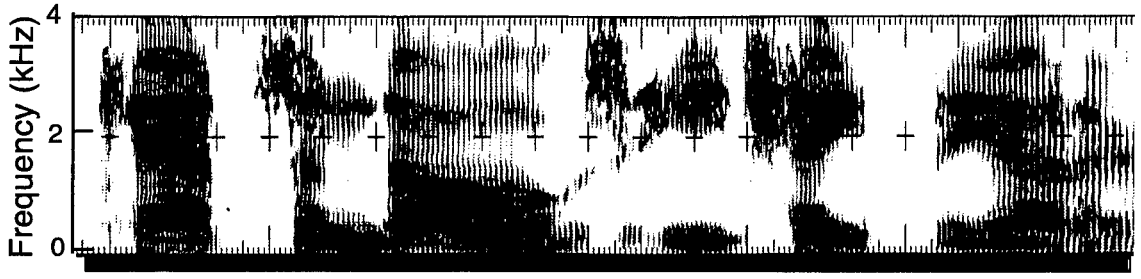
These transformations are unique and reversible. Some of their characteristics are:

- *Reflection Coefficients for LPC-10:* The LPC synthesis filter is a positive feedback filter. Thus, it becomes unstable if the filter has roots with a magnitude greater than unity. If PCs are transmitted, the stability must be ascertained for each frame. The use of RCs has an advantage because the synthesis filter never becomes unstable if the magnitude of each RC is less than unity. In fact, LPC-10 does not allow decoding of RCs that contribute to instability of the speech synthesizer.

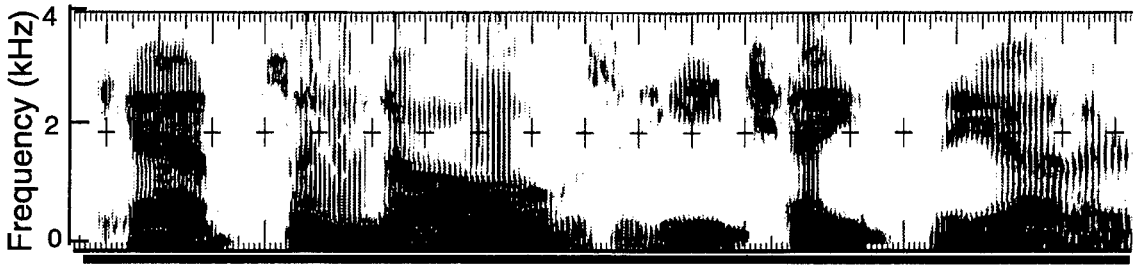
- *Line Spectrum Pairs for MELP*: LSPs are frequency domain parameters, and an error in an LSP only affects the speech spectrum near that frequency [6]. Since LSP errors are frequency selective, they can be quantized efficiently by exploiting the human perception characteristics. For example, since human perception is more tolerant to high-frequency errors, high frequency LSPs may be quantized more coarsely.

Speech Spectra With Preemphasis/Deemphasis Mismatch

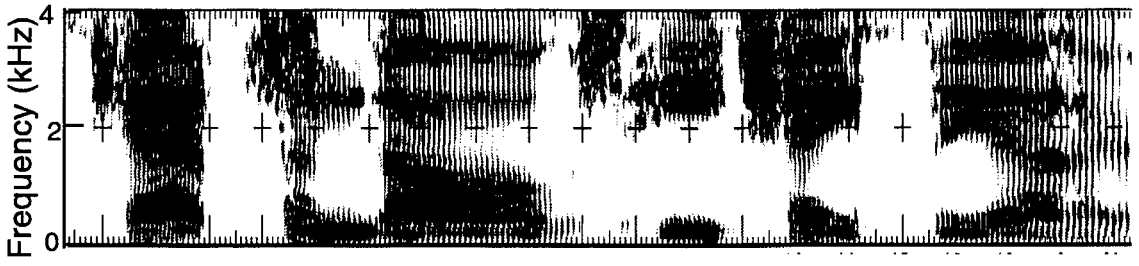
Because perfect compensation of preemphasis is computationally involved, it is tempting to skip the preemphasis compensation process such as not removing the preemphasis effect in filter coefficients when LPC-10 interoperates with MELP, or not introducing the preemphasis effect in filter coefficients when MELP interoperates with LPC-10. The result is substantial spectral distortions in the synthesized speech, making speech less intelligible. Figure 16 illustrates the ill effects.



(a) Ideal Case. This is the case where the LPC-10 or MELP transmitter interoperates with the receiver of its own kind, or the preemphasis effect is compensated.



(b) MELP into LPC-10 Without Preemphasis Compensation. Since there is no preemphasis in the MELP transmitter and there is deemphasis in the LPC-10 receiver, high-frequency components are attenuated in the synthesized speech. Speech does not sound too intelligible, particularly when heard in a noisy environment.



(c) LPC-10 into MELP Without Preemphasis Compensation. Since there is preemphasis in the LPC-10 transmitter and there is no deemphasis in then MELP receiver, high frequency components of the synthesized speech are boosted. Speech intelligibility is not affected due to strong high frequencies, but those high-passed speech spectra are not well encoded by MELP.

Fig. 16 — Spectral examples when preemphasis is not compensated in filter coefficients

Preemphasis Compensation in Filter Coefficients

In transcoding of MELP parameters to LPC-10 parameters (and vice versa), the most critical process is the introduction or removal of the preemphasis effect in filter coefficients. (See Fig. 7 to see why the preemphasis effect must be compensated in filter coefficients.) If this process is improperly implemented, speech is degraded significantly. Because the speech synthesizer in either MELP and LPC-10 is an all-pole filter, we have to ensure that filter coefficients will not cause filter instability. Introduction or removal of the preemphasis effect should not inadvertently cause the filter to become unstable. If an instability occurs, the synthesized speech is plagued by loud pops and other undesirable sounds.

First, we have to make an important decision as to which filter coefficients (among PCs, RCs and LSPs) are best suited to have preemphasis introduced, which is normally introduced in the speech waveform prior to the LPC analysis. Likewise, we have to make a decision as to which filter coefficients are best suited to have preemphasis nullified, which is normally performed in the synthesized speech waveform. Although parameter conversion requires computation time, it is not a serious drawback because parameter conversion is needed only once per frame (not sample by sample).

Among the three parameters (PCs, RCs and LSPs), we must decide which filter coefficients are most convenient to introduce or remove the preemphasis effect. We will analyze all three filter coefficients.

Use of Prediction Coefficients

The LPC analysis/synthesis process is often described in terms of PCs. Therefore, it appears to be convenient to use PCs to introduce or remove the preemphasis effect. But manipulating PCs is dangerous because the speech synthesizer may become unstable. Furthermore, a perfect compensation of preemphasis requires more than 10 coefficients, which is not permissible. To show this, we consider the transfer function of the MELP synthesis filter.

$$H_{MELP}(z) = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} \ \ldots \ - \alpha_{10} z^{-10}}, \tag{11}$$

where $\{\alpha\}$ is un-preemphasized PCs. On the other hand, the LPC-10 synthesis filter is

$$H_{LPC-10}(z) = \left[\frac{1}{1 - \beta_1 z^{-1} - \beta_2 z^{-2} \ \ldots \ - \beta_{10} z^{-10}}\right]\left[\frac{1}{1 - .9375 z^{-1}}\right], \tag{12}$$

where βs are PCs with preemphasis. Eq. (11) represents a 10-tap all-pole filter, whereas Eq. (12) represents an 11-tap all-pole filter. We cannot convert a 10-pole filter to an 11-pole filter, and vice versa. Therefore, PCs are not suited for transcoding.

Use of Reflection Coefficients

One advantage of using RCs over PCs is that the stability of the speech synthesizer is easily checked by the magnitude of the RCs. If the magnitude of each RC is less than unity, the synthesis filter is stable. But, as in the case of using PCs, more than 10 RCs are required to introduce or remove the preemphasis effect. Therefore, RCs are unsuited for transcoding.

Use of Line Spectrum Pairs

An advantage of using LSPs, similar to the use of RCs, is that the stability of the speech synthesis filter is easily checked. The speech synthesis filter is stable if the following conditions are met: (1) all LSP frequencies are naturally ordered (i.e., the first LSP is the lowest frequency and the succeeding LSPs are increasingly higher frequencies) and (2) the minimum distance from its neighboring LSP must be greater than approximately 50 Hz.

Not only is the filter stability easy to check, but the preemphasis effect may be easily introduced or removed in LSPs because of the following properties of LSPs:

- *Preemphasis upshifts LSPs*: If the speech waveform is preemphasized prior to the LPC analysis, estimated LSPs are of higher frequencies in comparison with those generated by the un-preemphasized speech waveform (Fig. 17).
- *Deemphasis downshifts LSPs:* If the speech waveform is un-preemphasized prior to the LPC analysis, estimated LSPs are of lower frequencies in comparison with those generated by the preemphasized speech waveform (see Fig. 17).

In other words, by upshifting LSPs, we can introduce the preemphasis effect in LSPs. Conversely, by downshifting LSPs, we can remove the preemphasis effect in LSPs. The magnitude of upshift is dependent on the speech as well as the preemphasis filter. Therefore, we must readjust LSPs when LPC-10 is interoperating with MELP, or vice versa.

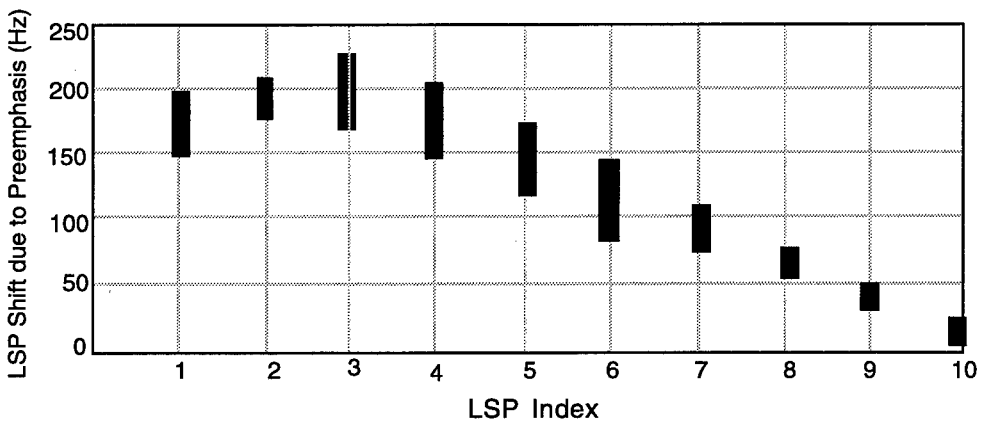


Fig. 17 — LSP shift caused by preemphasis. This figure indicates that the preemphasised speech waveform produces higher LSPs than non-preemphasized speech waveform. This figure is obtained from the analysis of a 3-min speech, and the preemphasis filter is as defined in Eq. (1).

Transcoding from LPC-10 Filter Coefficients to MELP Filter Coefficients

Four steps are involved in transcoding filter coefficients from LPC-10 to MELP, as indicated in Fig. 18. As in the transcoding of rms, steps 1 and 4 need no further elaboration because encoding and decoding rules for both LPC-10 and MELP are well defined and implemented in the hardware/software.

Step 2: RC-to-PC Conversion

The RC-to-PC conversion was previously explained in Eqs. (3) and (4) in connection with the rms transcoding.

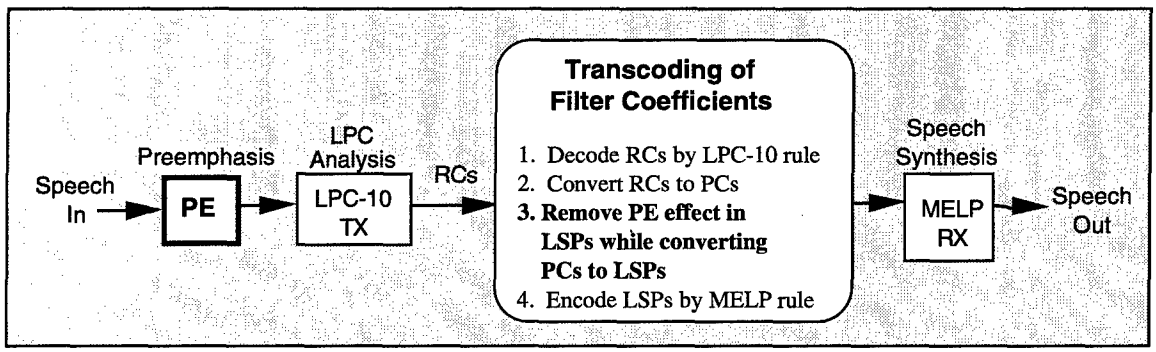


Fig. 18 — Filter coefficient transcoding from LPC-10 to MELP. The most critical step is removing the preemphasis effect (PE) in filter coefficients.

Step 3: PC-to-LSP Conversion

PCs from LPC-10 (generated from the preemphased speech waveform) must be converted to LSPs, which have the preemphasis effect removed. To accomplish these two objectives simultaneously, we use a different LSP estimation algorithm from what has been in some of the government-standard vocoders.

The PC-to-LSP conversion begins with the basic LPC equation, which related the input speech waveform to the prediction residual.

$$A(z) = 1 - \beta_1 z^{-1} - \beta_2 z^{-2} \ \ldots \ - \beta_{10} z^{-10}\,, \tag{13}$$

where βs are PCs with preemphasis (i.e., from LPC-10). The quantity z is a complex operator, which is defined as EXP(-jωτ), where ω is frequency and τ is the speech sampling time interval. To derive LSPs, $A(z)$ is decomposed into even and odd functions, denoted by $P(z)$ and $Q(z)$, respectively.

$$A(z) = \frac{1}{2}\left[P(z) + Q(z)\right], \tag{14}$$

where

$$\begin{aligned} P(z) &= A(z) + z^{-(n+1)} A(z^{-1}) \\ &= A(z)\left[1 + \frac{z^{-(n+1)} A(z^{-1})}{A(z)}\right] \end{aligned} \tag{15}$$

and

$$\begin{aligned} Q(z) &= A(z) - z^{-(n+1)} A(z^{-1}) \\ &= A(z)\left[1 - \frac{z^{-(n+1)} A(z^{-1})}{A(z)}\right], \end{aligned} \tag{16}$$

where n is the order of the LPC analysis system (in our case n=10). No information is lost in this even- and odd-decomposition because $A(z)$ can be reconstructed exactly using $P(z)$ and $Q(z)$ through the use of

Eq. (14). LSPs are the roots of $P(z)$ and $Q(z)$. In other words, LSPs are the frequencies that make the magnitude of either $P(z)$ or $Q(z)$ vanish.

In Eqs (15) and (16), the second term inside the bracket is an all-pass filter; that is, the amplitude response is independent of frequency and the phase response is a monotonically decreasing function of frequency. Let this all-pass filter be denoted by $R(z)$:

$$R(z) = \frac{z^{-(n+1)}A(z^{-1})}{A(z)}. \tag{17}$$

The phase response of $R(z)$ is

$$\varphi(kf_s) = -(n+1)(2\pi k_s t_s) - 2\tan^{-1}\left[\frac{\sum_{i=1}^{n}\alpha_i \sin(2\pi i k f_s t_s)}{1-\sum_{i=1}^{n}\alpha_i \cos(2\pi i k f_s t_s)}\right] \quad \text{k=1,2, ... ,} \tag{18}$$

where α_i is the ith prediction coefficient appearing in Eq. (11), t_s is the speech sampling time interval (125 μs in our case), and f_s is frequency for which the phase angle is computed. LSPs are the frequencies kf_s that make $\varphi(kf_s)$ either $-\pi$ or -2π radians. Eq. (18) is for deriving LSPs from the given speech waveform (i.e., a normal way of computing LSPs in the absence of a mismatch in preemphasis and deemphasis).

For transcoding of filter coefficients from LPC-10 to MELP, however, we have to remove the preemphasis effect in LSP. In effect, we have to reformulate Eq. (18) as if we have a front-end filter (a deemphasis filter in the present case). What we would like to know is the resultant phase in $R(z)$, if we have a front-end filter. Conversely, if we introduce the same amount of phase shift in $R(z)$ while we are estimating LSPs, then those LSPs will have the effect of the front-end filter (the deemphasis filter in the present case). Thus, let us introduce a deemphasis filter $H(z)$ in the all-pass filter $R(z)$, denoted by $R_1(z)$:

$$R_1(z) = \frac{z^{-(n+1)}A(z^{-1})}{A(z)}\left[\frac{H(z^{-1})}{H(z)}\right], \tag{19}$$

where $H(z)$ in this case is the deemphasis filter defined in Eq. (2).

$$\begin{aligned} H(z) &= H_{DE}(z) \\ &= \frac{1}{1-0.9375z^{-1}}. \end{aligned}$$

The phase response of the deemphasis filter, obtained from Eq. (2), is

$$\varphi_{DE}(kf_s) = -\tan^{-1}\left[\frac{-0.9375\sin(2\pi k_s t_s)}{1-0.9375\cos(2\pi k_s t_s)}\right]. \tag{20}$$

The phase response of the deemphasis filter was previously plotted in Fig. 6. Combining Eqs. (20) and (18) gives the phase response of $R(z)$ with the deemphasis filter. Thus,

$$\varphi_1(kf_s) = -(n+1)(2\pi k_s t_s) - 2\tan^{-1}\left[\frac{\sum_{i=1}^{n}\alpha_i \sin(2\pi ikf_s t_s)}{1-\sum_{i=1}^{n}\alpha_i \cos(2\pi ikf_s t_s)}\right] - 2\tan^{-1}\left[\frac{-0.9375\sin(2\pi kf_s t_s)}{1-.9375\cos(2\pi kf_s t_s)}\right], \quad (21)$$

where k = 1, 2, ... Again, LSPs are the frequencies that make the phase angles of $\varphi_1(kf_s)$ either -π or -2π.

Demonstration of MELP Filter Coefficients Transcoded from LPC-10 Filter Coefficients

Figure 19 shows an example of three spectra estimated by LPC-10, MELP and transcoding. In each case, LSPs are computed, and they are represented in terms of the amplitude spectra to make an easy comparison.

1. *The original LPC-10 spectrum computed from the preemphasized speech waveform*: Eq. (18) is used to estimate LSPs. Using these LSPs, the speech spectral envelope is computed. This is a reference spectrum for comparison (thin line).

2. *The original MELP spectrum computed from the non-preemphasized speech waveform*: Eq. (18) is used to estimate LSPs. Using these LSPs, the speech spectral envelope is computed. This is also a reference spectrum for comparison (thick line).

3. *The MELP spectrum computed from the transcoded LPC-10 filter coefficients*: Eq. (21) is used to compute LSPs (crossmarks on or near thick line).

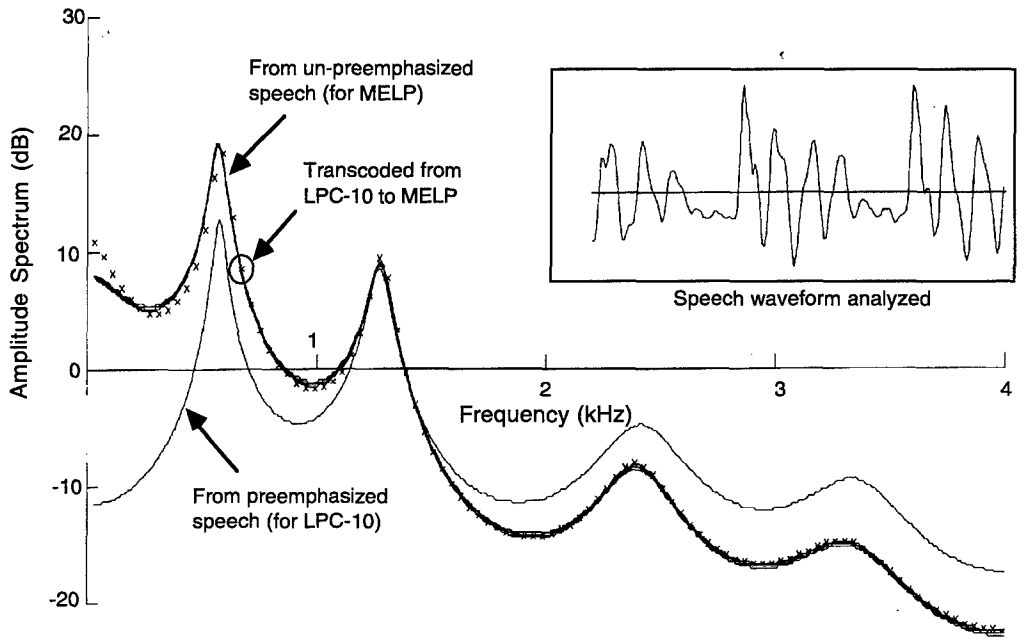


Fig. 19 — Speech spectra estimated by LPC-10, MELP, and via transcoding from LPC-10 to MELP. The transcoded MELP spectrum agrees with the original MELP spectrum very well. Small discrepancies below approximately 100 Hz are not audible.

Transcoding from MELP Filter Coefficients to LPC-10 Filter Coefficients

Five steps are involved in transcoding filter coefficients from MELP to LPC-10, as indicated in Fig. 20. As in preceding cases, steps 1 and 5 need no further elaboration because the encoding and decoding rules for both LPC-10 and MELP are well defined and have been implemented in LPC-10 and MELP.

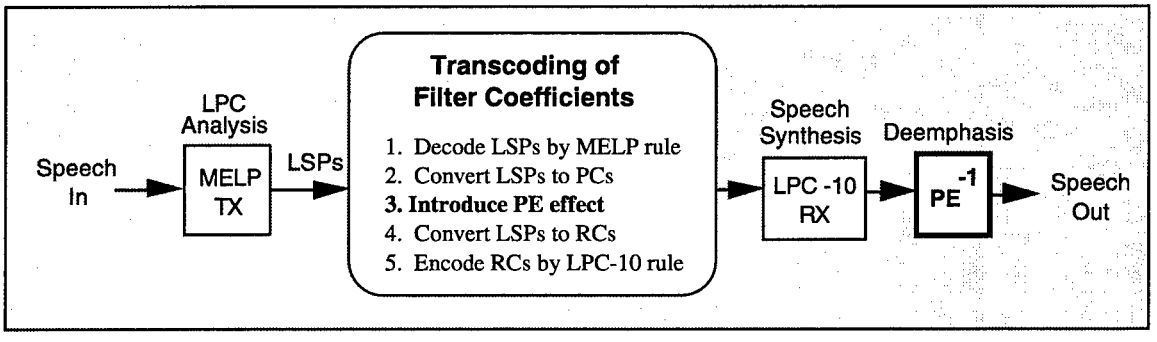


Fig. 20 — Filter coefficient transcoding from MELP to LPC-10. The most critical step is introducing the preemphasis effect (PE) in filter coefficients.

Step 2: Convert LSPs to PCs

Referring to an even and odd decomposition of $A(z)$ discussed in connection with Eqs. (13) through (16), LSPs are roots of $P(z)$ and $Q(z)$ along the unit circle of the complex z plane. Thus, $P(z)$ may be expressed in terms of those roots.

$$P(z) = (1 + z^{-1}) \prod_{k=1}^{5} (1 - \varepsilon^{j\theta_k} z^{-1})(1 - \varepsilon^{-j\theta_k} z^{-1}), \tag{22}$$

where θ_k is the location of the lower frequency of the kth LSP. If a line-spectrum frequency is 0 Hz, then $\theta_k = 0$ radian; if a line-spectrum frequency is 4 kHz (half sampling frequency), then $\theta_k = \pi$ radians. The root at $z = -1$ is an artifact generated during the even and odd decomposition. It is time-invariant, and it contains no speech information.

Likewise, the transfer function of the difference filter is

$$Q(z) = (1 - z^{-1}) \prod_{k=1}^{5} (1 - \varepsilon^{j\theta_k'} z^{-1})(1 - \varepsilon^{-j\theta_k'} z^{-1}), \tag{23}$$

where θ_k' is the location of the upper frequency of the kth LSP. The root at $z = 1$ is a byproduct of the even and odd decomposition, and it contains no speech information.

From Eq. (14), the transfer function of the LPC analysis filter in terms of the even and odd filters is

$$A(z) = \frac{1}{2}[P(z) + Q(z)], \tag{24}$$

which is in the form of

$$A(z) = 1 + \mu_1 z^{-1} + \mu_2 z^{-2} + \ \ldots \ + \mu_{10} z^{-10}, \tag{25}$$

where μ's are new PCs of $A(z)$. Comparing Eq. (25) with Eq. (13) indicates that the kth PC is

$$PC(z) = -\mu_k . \tag{26}$$

Step 3: Introduce the Preemphasis Effect in LSPs

We use the identical technique used for transcoding from LPC-10 to MELP. We introduce the preemphasis effect in LSPs while we are computing LSPs. From Eq. (19),

$$R_1(z) = \frac{z^{-(n+1)}A(z^{-1})}{A(z)}\left[\frac{H(z^{-1})}{H(z)}\right],$$

where H(z), in this case, is the preemphasis filter defined in Eq. (1).

$$\begin{aligned} H(z) &= H_{PE}(z) \\ &= 1 - 0.9375z^{-1} \ . \end{aligned}$$

The phase response of the preemphasis filter, obtained from Eq. (1), is

$$\varphi_{PE}(kf_s) = \tan^{-1}\left[\frac{-0.9375\sin(2\pi k_s t_s)}{1-0.9375\cos(2\pi k_s t_s)}\right], \tag{27}$$

which was plotted earlier in Fig. 6.

This is the case where LSPs are computed while adding the preemphasis effect in LSPs (i.e., the case of MELP-to-LPC-10 transcoding). In this case, $H(z) = H_{PE}(z)$

$$\varphi_1(kf_s) = -(n+1)(2\pi k_s t_s) - 2\tan^{-1}\left[\frac{\sum_{i=1}^{n}\alpha_i \sin(2\pi i k f_s t_s)}{1-\sum_{i=1}^{n}\alpha_i \cos(2\pi i k f_s t_s)}\right] + 2\tan^{-1}\left[\frac{-0.9375\sin(2\pi k f_s t_s)}{1-.9375\cos(2\pi k f_s t_s)}\right], \tag{28}$$

where k = 1,2,... From Eq. (28), LSPs are the frequencies that make the phase angle $\varphi_1(kf_s)$ = -π or -2π. The third term in the right-hand member of Eq. (28) is the phase contributed by the preemphasis.

Step 4: Convert LSPs to RCs

LSPs are converted to RCs in two steps. First, convert the LSPs to PCs by the method discussed previously in Eqs. (22) through (26), then convert the resultant PCs to RCs. From Eqs. (3) and (4), the PCs in terms of RCs are:

$$\beta_{j|n+1} = \beta_{j|n} - k_{n+1}\beta_{n+1-j|n} \qquad \text{j = 1, 2, ..., n} \tag{3}$$

with

$$\beta_{n+1|n+1} = k_{n+1}, \tag{4}$$

where $\beta_{j|n+1}$ means the jth prediction coefficient (with preemphasis) in the $(n+1)^{th}$ iteration. Let j be replaced by $n+1-j$ in Eq. (3). Thus,

$$\alpha_{n+1-j|n+1} = \alpha_{n+1-j|n} - k_{n+1}\alpha_{j|n}. \tag{29}$$

Eqs. (3) and (29) are a set of simultaneous equations with two unknowns, $\alpha_{n+1-j|n}$ and $\alpha_{j|n}$. Solving for $\alpha_{j|n}$, or alternatively for $\alpha_{n+1-j|n}$, gives

$$\alpha_{j|n} = \frac{\alpha_{j|n+1} + k_{n+1}\alpha_{n+1-j|n+1}}{1-k_{n+1}^2}, \tag{30}$$

where $j = 1, 2, 3, ..., n$. Eq. (30) converts a set of PCs to a set of RCs, where $k_n = \alpha_{n/n}$.

Demonstration of LPC-10 Filter Coefficient Transcoded from MELP Filter Coefficients

Figure 21 shows an example of three spectra estimated by LPC-10, MELP, and transcoding. In each case, LSPs are computed, and they are represented in terms of the amplitude spectra to make an easy comparison.

1. *The original LPC-10 spectrum computed from the preemphasized speech waveform*: Eq. (18) is used to estimate LSPs. Using these LSPs, the speech spectral envelope is computed. This is a reference spectrum for comparison (thin line).

2. *The original MELP spectrum computed from the non-preemphasized speech waveform*: Eq. (18) is used to estimate LSPs. Using these LSPs, the speech spectral envelope is computed. This is also a reference spectrum for comparison (thick line).

3. *The LPC-10 spectrum computed from the transcoded MELP filter coefficients*: Eq. (28) is used to compute LSPs (crossmarks on or near thin line).

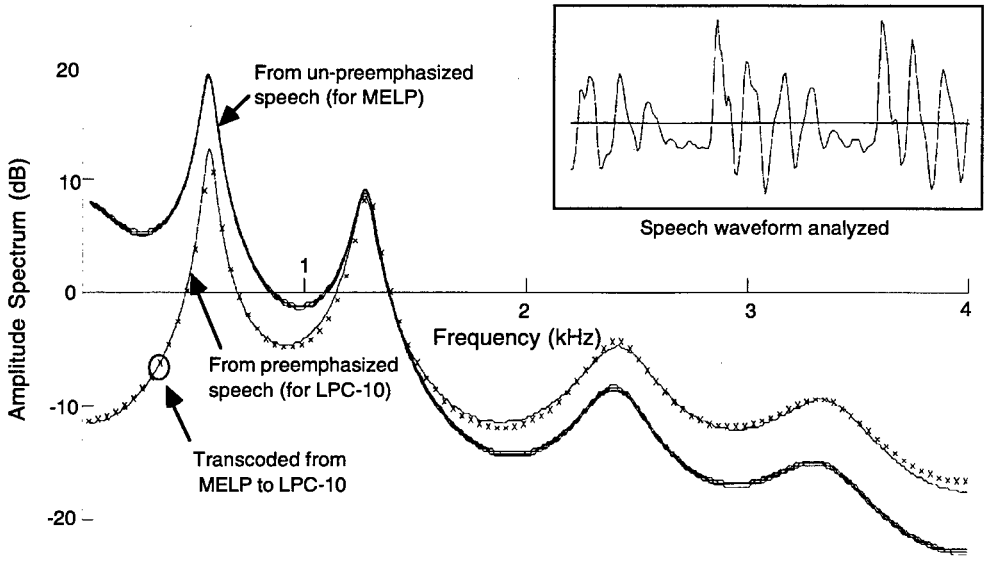


Fig. 21 — Speech spectra estimated by MELP, LPC-10, and MELP-to-LPC-10 transcoding. As noted, the spectrum computed from the MELP-to-LPC-10 transcoded LSPs is as good as the spectrum computed from the original MELP LSPs.

TRANSCODING OF EXCITATION PARAMETERS

Both LPC-10 and MELP transmit excitation parameters that control characteristics of the excitation signal. Therefore, these parameters must also be transcoded. Transcoding is based on rules, rather than computations, as in the transcoding of the speech rms value or filter coefficients.

Background

Both LPC-10 and MELP have 54 bits to encode speech data at a frame rate of 44.44 Hz (Table 1). These 54 bits are divided to encode individual speech parameters including speech rms, filter coefficients, and excitation parameters. Among these parameters, filter coefficients require the greatest number of bits because they represent a complex speech spectral envelope. LPC-10 uses as much as 41 bits (approximately 76% of 54 bits) to encode 10 RCs. LSPs, however, do not require as many bits as RCs. MELP capitalized on this technology to use only 25 bits (46% of the 54 bits) to encode LSPs. Therefore, MELP has more bits available to encode excitation parameters than LPC-10.

Table 1 — Bit Allocations for LPC-10 and MELP

	Voiced Speech		**Unvoiced Speech**	
	LPC-10	MELP	LPC-10	MELP
Filter Coefficients	41 bit(s)	25 bit(s)	21 bit(s)	25 bit(s)
Speech Rms	5	8	5	8
Excitation Signal				
Pitch and Overall Voicing	7	7	7	7
Bandpass Voicing	0	4	0	0
Fourier Magnitudes	0	8	0	0
Aperiodic Flag	0	1	0	0
Error Protection	0	0	20	13
Synchronization	1	1	1	1
TOTAL	54	54	54	54

Voiced Excitation Parameters:

The speech waveform of voiced speech (vowels) is more complex than that of the unvoiced speech. Likewise, the voiced speech spectrum is more complex than the unvoiced speech spectrum. Furthermore, the human ear is rather sensitive to misplaced resonant frequencies or spectral flutters caused by coarse filter coefficient quantization. Therefore, LPC-10 or MELP spends the entire 54 bits (less one sync bit) per frame to encode speech parameters. There are four voiced excitation parameters involved in transcoding.

- *Pitch and Overall Voicing:* The pitch parameter controls the fundamental pitch frequency of the synthesized speech. Pitch is semilogarithmically quantized from approximately 50 to 400 Hz into a 6-bit quantity for both LPC-10 and MELP. In addition, one bit is allocated to represent the overall voicing decision. Although MELP quantizes pitch slightly differently than LPC-10, the respective decoding tables provide an appropriate pitch value for transcoding. Thus, the pitch and overall voicing parameter, either for the voiced or unvoiced frame, is directly transcodable.

- *Bandpass Voicing:* MELP makes additional voicing decisions for each 800 Hz sub-band starting from 800 Hz to 4 kHz. Therefore, 4 bits are required for the sub-band voicing decision. As a result, the excitation signal for MELP is a mixture of periodic (a pulse train) and aperiodic (random noise) signals where mixing characteristics are controlled by the 4 bits. On the other hand, LPC-10 does not make bandpass voicing. If speech is voiced, however, the low-passed periodic component and high-passed aperiodic components are mixed. There are no control parameters to adjust mixing characteristics.

- *Fourier Magnitudes*: The prediction residual, the ideal excitation signal for the LPC-based system, in general, does not have a flat spectral envelope. There is always some amount of remnant resonant frequencies and other coloration. MELP transmits 10 magnitudes of the residual spectrum sampled at pitch harmonics encoded at 8 bits. LPC-10 spectrally shapes the excitation signal by a filter that introduces weak resonant frequencies. The filter has the roots similar to those of the speech synthesizer but reduced magnitudes.

- *Aperiodic Flag:* The natural speech waveform has pitch jitters. MELP allocates 1 bit to make a binary decision to indicate the presence of substantial pitch jitters in the voiced speech. LPC-10 does not generate such information.

Unvoiced Excitation Parameters:

The speech waveform (or spectrum) of unvoiced speech (consonants) is not as complex as that of vowels. Moreover, the human ear is much more tolerant to the variability of consonants (i.e., /s/, /ch/. /ch/, /t/, etc.) caused by coarse parameter quantization. Therefore, both LPC-10 and MELP encode only one excitation parameter for generating unvoiced speech sounds, which is pitch and overall voicing. Furthermore, both LPC-10 and MELP use 7 bits to encode this parameter. Thus, the unvoiced excitation parameter is directly transcodable.

Transcoding Rules

The transcoding rules for excitation signal parameters are simple because LPC-10 does not encode three of the four parameters that MELP encodes. As listed in Table 2, pitch and overall voicing is the only parameter that is directly converted from LPC-10 to the MELP, and vice versa. The rest (bandpass voicing, Fourier magnitudes, and aperiodic flag) are fixed values.

Table 2 — Transcoding Rules for Excitation Parameters

	LPC-10 to MELP	MELP to LPC-10
Pitch and Overall Voicing	Directly transcodable	
Bandpass Voicing	All 4 bandpass voicings are made equal to overall voicing	Discard
Fourier Magnitudes	All 10 magnitudes set to unity	Discard
Aperiodic Flag	Information not available	Discard

INTELLIGIBILITY TESTS

In previous sections, we have demonstrated that the transcoding algorithms provided accurate converted values for both the speech rms value and filter coefficients. Since this is true, the speech intelligibility of LPC-10 and MELP interoperating together should be at least as good as the intelligibility of LPC-10, which is the weaker voice encoder of the two in series. To confirm this hypothesis, we performed the Diagnostic Rhyme Test (DRT), a formalized test to evaluate the initial consonant intelligibility, using male speakers speaking in a quiet environment. The result is summarized in Fig. 22.

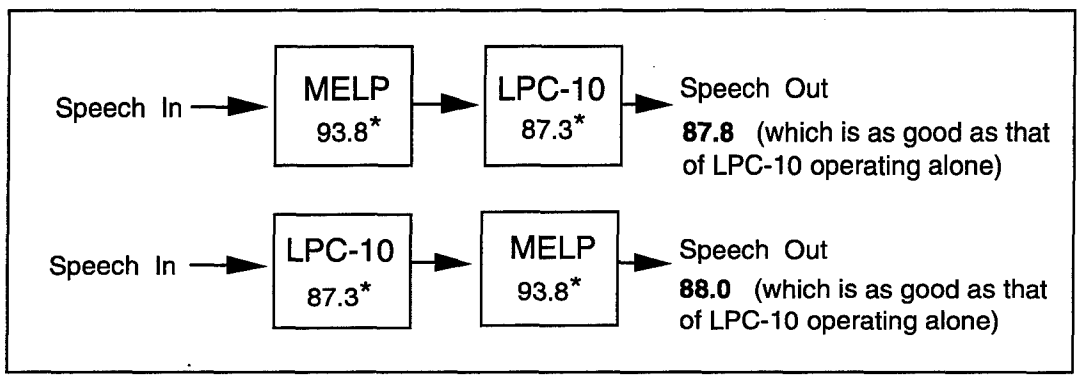


Fig. 22 — DRT scores when LPC-10 is interoperating with MELP using the transcoding algorithms presented in this report. The DRT scores of the individual voice encoder are indicated by asterisks. This figure indicates that the DRT scores of the transcoded speech are as as good as LPC-10, which is the weaker voice encoder. In other words, transcoding does not degrade speech intelligibility in comparison with the weak vocoder in the link.

CONCLUSIONS

This report presents an effective way of interoperating between two different voice encoders by converting speech parameters directly from one voice encoder to another. In this way, we can bypass all of the following functional operations that are known to degrade speech — the digital-to-analog converter, reconstruction filter, speech synthesis, analog-to-digital converter, anti-aliasing filter, and speech analysis. Recently, someone coined the word *transcoding* for this form of direct interoperation. Yet, no one ever investigated any aspect of transcoding before. This report marks the first of transcoding investigation, specifically for two DoD narrowband voice algorithms (LPC-10 and MELP).

This R&D effort is motivated by the fact that we need a technology that enables the old DoD narrowband voice algorithm (LPC-10) featured in 40,000 presently deployed ANDVTs to interoperate directly with the future DoD narrowband voice algorithm (MELP) without hurting speech intelligibility. The technology developed in this report improves the connectivity and speech quality of DoD narrowband secure voice systems.

ACKNOWLEDGMENTS

The authors thank CAPT Galik of SPAWAR PMW161 and Chris Newborn and Mike Weber of SPAWAR PMW161-3 for supporting this R&D effort. We worked closely with the NRL Communication Security Group in defining Navy problems and transferring the technologies developed at NRL to the Navy Fleet. We thank Stan Chincheck, John Gessner, and Gautum Trivedi who made this coordinated effort effective.

REFERENCES

1. "Performance Specifications for LPC Processor," TT-BI-4210-0087 A, Joint Tactical Communications Office, Fort Monmouth, NJ, 1980.

2. "FNBDT -210 (Revision 1.0)," An internal document of the National Security Agency, 9800 Savage Road, Fort George G. Meade, MD 20755, 1998.

3. Alan McCree, et al., "A 2.4 kb/s MELP Coder for the New U.S. Federal Standard," *Proc. 1996 IEEE Int. Conf. Acoust., Speech, and Signal Processing,* 1996, pp. 200-203.

4. "Analog to Digital Conversion of Voice by 2,400 Bits/Second Linear Predictive Coding," Federal Standard 1015, General Services Administration, Specification Unit, Washington, DC 20407, 1984.

5. S.A. Tretter, *Introduction to Discrete-Time Signal Processing* (John Wiley & Sons, New York, NY, 1976).

6. G.S. Kang and L.J. Fransen, "Low-Bit Rate Speech Encoders Based on Line-Spectrum Frequencies (LSFs)," NRL Report 8857 (1985).